DOCUMENT RESUME

ED 058 276   TM 000 989

AUTHOR   Lord, Frederic M.
TITLE   A Numerical Empirical Bayes Procedure for Finding an
                Interval Estimate.
INSTITUTION   Educational Testing Service, Princeton, N.J.
SPONS AGENCY   Office of Naval Research, Washington, D.C. Personnel
                and Training Research Programs Office.
REPORT NO   RB-71-46
PUB DATE   Jul 71
NOTE   26p.

EDRS PRICE   MF-$0.65 HC-$3.29
DESCRIPTORS   *Bayesian Statistics; Computer Programs; Expectation;
                Goodness of Fit; *Mathematical Applications;
                Mathematical Models; Multiple Regression Analysis;
                Predictive Measurement; Probability Theory;
                *Sampling; *Statistics
IDENTIFIERS   *Statistical Inference

ABSTRACT
                A numerical procedure is outlined for obtaining an
interval estimate of a parameter in an empirical Bayes estimation
problem. The case where each observed value x has a binomial
distribution, conditional on a parameter zeta, is the only case
considered. For each x, the parameter estimated is the expected value
of zeta given x. The main purpose is to throw some light on the
effectiveness of empirical Bayes estimation in samples of various
sizes. Illustrative numerical results are presented. (Author)

RB-71-46

A NUMERICAL EMPIRICAL BAYES PROCEDURE

FOR FINDING AN INTERVAL ESTIMATE

Frederic M. Lord

Educational Testing Service

Princeton, New Jersey

July 1971

A NUMERICAL EMPIRICAL BAYES PROCEDURE

FOR FINDING AN INTERVAL ESTIMATE

Frederic M. Lord

Educational Testing Service

Princeton, New Jersey

July 1971

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*	2a. REPORT SECURITY CLASSIFICATION
Educational Testing Service Princeton, New Jersey 08540	Unclassified
	2b. GROUP

3. REPORT TITLE

A NUMERICAL EMPIRICAL BAYES PROCEDURE FOR FINDING AN INTERVAL ESTIMATE

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*
Technical Report

5. AUTHOR(S) *(First name, middle initial, last name)*

Frederic M. Lord

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
July 1971	18	11

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
N00014-69-C-0017	
b. PROJECT NO. NR No. 150-303	RB-71-46
c.	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
d.	

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Personnel and Training Research Programs Office of Naval Research Arlington, Virginia 22217

13. ABSTRACT

A numerical procedure is outlined for obtaining an interval estimate of a parameter in an empirical Bayes estimation problem. The case where each observed value x has a binomial distribution, conditional on a parameter ζ , is the only case considered. For each x , the parameter estimated is the expected value of ζ given x . The main purpose is to throw some light on the effectiveness of empirical Bayes estimation in samples of various sizes. Illustrative numerical results are presented.

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Empirical Bayes						
Statistical Inference						
Confidence Intervals						
Statistical Estimation						
Compound Distributions						
Mixtures of Distributions						
Binomial Distribution						
Compound Binomial Distribution						

A NUMERICAL EMPIRICAL BAYES PROCEDURE FOR FINDING AN INTERVAL ESTIMATE

Summary

A numerical procedure is outlined for obtaining an interval estimate of a parameter in an empirical Bayes estimation problem. The case where each observed value x has a binomial distribution, conditional on a parameter ζ, is the only case considered. For each x, the parameter estimated is the expected value of ζ given x. The main purpose is to throw some light on the effectiveness of empirical Bayes estimation in samples of various sizes. Illustrative numerical results are presented.

1. Introductory Numerical Example

The following example illustrates a type of situation in which the usual estimator for a binomial parameter can be very misleading. A sample of $n = 10$ independent observations was drawn from a Bernouilli distribution with probability of success π. This was repeated for $N = 10,000$ different values of π, these being obtained by random sampling from a prechosen distribution of π's. The sample proportion p of successes was found for each of the 10,000 samples. The resulting sample distribution $Nf(p)$ of the 10,000 values of p is shown below:

p	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
$Nf(p)$	7	102	477	1140	2053	2476	2027	1175	437	98	10
$(\underline{\pi}, \bar{\pi})$	.189-.505		.469-.505		.490-.505		.493-.509		.493-.527		.493-.840

The usual estimator of the value of π giving rise to any particular sample of 10 is the observed p for that sample, this being a sufficient

statistic if the sample of 10 is considered by itself. For even values of 10p , the table also shows interval estimates $(\underline{\pi}, \bar{\pi})$ of the regression of π on p , obtained by the empirical Bayes approach to be described in this paper.

The purpose of the above numerical example is to display the disagreement between the usual estimates p of π and the empirical Bayes interval estimates shown. The reason for such sharp disagreement is that the population distribution of π chosen to generate these artificial data, until now concealed from the reader, actually had a negligible standard deviation so that $Nf(p)$ is in effect a random sample of 10,000 from a binomial distribution with $\pi = 0.5$. Although this information was not available for estimation purposes, the empirical Bayes estimation procedure partially recovered equivalent information from the sample $Nf(p)$. A pleasing result is that each interval estimate here turns out to include the true value $\pi = 0.5$.

2. Mathematical Formulation

Suppose that for each randomly drawn observation, there is not only an observed value x of the discrete random variable X , but also a corresponding unobservable value ζ of the continuous random variable Z . For any given ζ , x is drawn at random from the binomial distribution

$$h(x|\zeta) \equiv \text{Prob}(X = x|\zeta) = \binom{n}{x}\zeta^x(1 - \zeta)^{n-x} \quad , \qquad x = 0,1,\ldots, n , \quad (1)$$

where n is known. Given randomly drawn observations $x_1, x_2, \ldots, x_N$, the

ζ corresponding to any particular observation x_a can be estimated by empirical Bayes point-estimation methods (for example, Robbins, 1956; Maritz, 1966; Copas, 1969; Griffin & Krutchkoff, 1971). More needs to be known about the effectiveness of such methods where n and N are not both very large.

Denote by $G(\zeta)$ the unknown cumulative distribution function of Z for the population from which the unobservable values are drawn. In view of (1), it will be assumed that the range of ζ is $0 \leq \zeta \leq 1$. Ordinarily, $G(\zeta)$ is thought of as continuous, but we will not exclude the possibility that it is a step function.

The (unconditional) probability distribution of X for the population can be written

$$\phi_G(x) = \int_0^1 h(x|\zeta)\, dG(\zeta) \quad , \qquad x = 0, 1, \ldots, n \quad . \tag{2}$$

The observed sample distribution of $x_1, x_2, \ldots, x_N$, to be denoted by $f(x)$, is the distribution of a random sample from $\phi_G(x)$.

If $G(\zeta)$ were known, the value of ζ corresponding to any observed x would in common practice be estimated by the regression function

$$\mu_{Z|x} \equiv \mathcal{E}_G(Z|x) = \frac{1}{\phi_G(x)} \int_0^1 \zeta\, h(x|\zeta)\, dG(\zeta) \quad , \quad x = 0, 1, \ldots, n \quad , \tag{3}$$

usually called the Bayes estimator. When $G(\zeta)$ is unknown, as here, the problem of finding a point estimator $\hat{\zeta}$ for the ζ corresponding to an observed x is a standard empirical Bayes problem. Typically, a point

estimate $(\hat{\mu}_{Z|x_a}$, say) of $\mathcal{E}(Z|x_a)$, the regression of Z on x evaluated at x_a , is used as the empirical Bayes estimator of ζ for observation a . The binomial case considered here is of particular interest (as compared, for example, to the Poisson) since in the binomial case $G(\zeta)$ is unidentifiable--complete knowledge of $\phi_G(x)$ is sufficient to determine only the first n moments of $G(\zeta)$ (Skellam, 1948).

The present note is explicitly concerned with an _interval_ estimator $(\underline{\mu}_{\alpha x}, \bar{\mu}_{\alpha x})$ of $\mu_{Z|x}$, $x = 0, 1, \ldots, n$, $G(\zeta)$ being unknown. Given the empirical Bayes model already specified and a sample with the observed distribution $f(x)$, what range of values for $\hat{\mu}_{Z|x}$, $x = 0, 1, \ldots, n$, is reasonably consistent with the observed $f(x)$?

Consider the set Γ_α consisting of all distribution functions $G(\zeta)$ such that the chi square (χ^2_G) between $\phi_G(x)$, defined by (2), and the given $f(x)$ is less than $\chi^2_{1-\alpha}$, the $1 - \alpha$ percentile of the chi square distribution:

$$\chi^2_G \equiv \sum_{x=0}^{n} \frac{[f(x) - \phi_G(x)]^2}{\phi_G(x)} \leq \chi^2_{1-\alpha} \quad . \tag{4}$$

For each possible value of x , find $\bar{\mu}_{\alpha x} \equiv \underset{\Gamma_\alpha}{\text{Max}}(\mu_{Z|x})$ and $\underline{\mu}_{\alpha x} \equiv \underset{\Gamma_\alpha}{\text{Min}}(\mu_{Z|x})$, the maximum and minimum values of (3) in Γ_α . For any given x , any value of $\mu_{Z|x}$ in the interval $(\underline{\mu}_{\alpha x}, \bar{\mu}_{\alpha x})$ is consistent with the data; values outside this interval will be considered implausible.

The foregoing is not an ideal way to set up an interval estimate. Until better methods are implemented, however, it can throw some light on the accuracy of empirical Bayes point estimation. For the standard empirical Bayes problem of estimating ζ , any point estimate $\hat{\zeta}$ in the interval

$(\underline{\mu}_{\alpha x}, \bar{\mu}_{\alpha x})$ would minimize the squared errors of estimation for some $G(\zeta)$ in Γ_α . Any estimate outside this interval would not.

If desired, the interval described above can be interpreted as a confidence interval for $\mu_{Z|x}$ with confidence level $> 1 - \alpha$. This may be seen as follows. For any given $G(\zeta)$, Γ_α is a random variable so defined as to include $G(\zeta)$ $1 - \alpha$ of the time. Thus $(\underline{\mu}_{\alpha x}, \bar{\mu}_{\alpha x})$ must include the true $\mu_{Z|x}$ at least $1 - \alpha$ of the time.

3. Bounds for the Regression of Z on x

Substituting (1) in (2) and expanding, we have

$$\phi(x) = \binom{n}{x} \int_0^1 \zeta^x (1 - \zeta)^{n-x} \, dG(\zeta)$$

$$= \binom{n}{x} \sum_{r=0}^{n-x} \binom{n-x}{r} (-1)^r \mu'_{r+x} , \qquad x = 0, 1, \ldots, n \quad , \quad (5)$$

where μ'_s is the moment about the origin of order s for the distribution $G(\zeta)$. Ruling out the degenerate case where Z takes no values other than 0 or 1 (and where consequently $\phi(x) = 0$ unless $x = 0$ or n), we have that $\mu'_s > 0$ for all s . Similarly, from (3),

$$\mu_{Z|x} = \frac{\displaystyle\sum_{r=0}^{n-x} \binom{n-x}{r} (-1)^r \mu'_{r+x+1}}{\displaystyle\sum_{r=0}^{n-x} \binom{n-x}{r} (-1)^r \mu'_{r+x}} = \frac{\displaystyle\sum_{R=x}^{n} \binom{n-x}{R-x} (-1)^{R-x} \mu'_{R+1}}{\displaystyle\sum_{R=x}^{n} \binom{n-x}{R-x} (-1)^{R-x} \mu'_{R}} , \quad x = 0, 1, \ldots, n \quad .$$

$$(6)$$

Consider first a restricted case where $\mu_1', \mu_2', \ldots, \mu_n'$ are fixed. In this special case, (6) is seen to be maximized (minimized) by maximizing (minimizing) μ_{n+1}' if $n - x$ is even, by minimizing (maximizing) μ_{n+1}' if $n - x$ is odd. A theorem of Markov (see Possé, 1886, sections V8 and V9; or Karlin & Shapley, 1953) shows that if $\mu_1', \mu_2', \ldots, \mu_n'$ (considered fixed) are the moments of some frequency distribution, then the maximum (minimum) value of μ_{n+1}' is uniquely attained when $G(\zeta)$ is a certain specified step function. If n is even, all the frequency is concentrated at exactly $n/2 + 1$ different values of ζ, including $\zeta = 1$ if μ_{n+1}' is maximized, $\zeta = 0$ if μ_{n+1}' is minimized. The situation for odd n is similar, but need not be detailed here.

This leads to the key conclusion that in order to find $G(\zeta)$ maximizing or minimizing (6) for fixed x, when $\mu_1', \mu_2', \ldots, \mu_n'$ are given , it is only necessary to determine $M \equiv n/2$ unknown values $\zeta_0, \zeta_1, \ldots, \zeta_{M-1}$ or $\zeta_1, \zeta_2, \ldots, \zeta_M$, together with the corresponding M unknown frequencies $g_0, g_1, \ldots, g_{M-1}$ or $g_1, g_2, \ldots, g_M$. It is not necessary to admit to consideration any of the continuous frequency distributions on $(0,1)$ nor any discrete distribution with more than M unknown values of ζ .

In our actual problem, we wish to find the probability distribution $\bar{g}_{\alpha x}(\zeta)$, say, maximizing (6) or the $\underline{g}_{\alpha x}(\zeta)$ minimizing (6) subject to the restriction that the corresponding cumulative distribution function is in Γ_α . This restriction does not change the key conclusion stated above, since holding $\mu_1', \mu_2', \ldots, \mu_n'$ fixed also holds $\phi_G(0), \phi_G(1), \ldots, \phi_G(n)$ fixed, because of (5), and thus fixes χ_G^2 . Thus the extremizing $G(\zeta)$ will still be a step function with exactly $M + 1$ different values of ζ , as before. (For example, let $\mu_1', \mu_2', \ldots, \mu_n'$ be the first n moments of

10

$g_{\alpha x}$. Markov's theorem states that the minimizing $G(\zeta)$ with these moments is of the special form described. Thus if $g_{\alpha x}$ exists, it is of this special form.)

It is, of course, always possible that Γ_α is empty. This situation has not yet arisen in practical application. The smaller the value of α chosen, the less likely it is that Γ_α will be empty.

For a discrete $G(\zeta)$ such as $g_{\alpha x}(\zeta)$ the first line of (5) can be written

$$\phi(x) = \binom{n}{x} \sum_{m=0}^{M} g_m \zeta_m^x (1 - \zeta_m)^{n-x} \quad , \tag{7}$$

where either $g_0 = 0$ or $g_M = 0$. Similarly, (3) becomes

$$\mu_Z|_x = \frac{\binom{n}{x}}{\phi(x)} \sum_{m=0}^{M} g_m \zeta_m^{x+1} (1 - \zeta_m)^{n-x} \quad . \tag{8}$$

Formulas derived by Markov (not given here) provide the explicit solution (if any) to the extremization problem when $\mu_1', \mu_2', \ldots, \mu_n'$ are fixed. These formulas do not help with the more general problem to be solved here, which seems to require the numerical methods of mathematical programming.

4. Numerical Methods

Note that the use of a step function for $G(\zeta)$ here is required by the problem itself, not imposed for the convenience of the writer. For simplicity, the following discussion deals only with minimization. Maximization is essentially similar.

11

Thanks to Markov, our problem is now to find ζ_m and $g_m{'}$,
$m = 0,1,\ldots,M - 1$ or $m = 1,2,\ldots,M$, so as to minimize (8), subject to
the restrictions

$$0 \leq \zeta_m \leq 1 , \quad g_m \geq 0 , \text{ and } \Sigma g_m \leq 1 , \tag{9}$$

and also subject to (4). The problem thus stated can be solved numerically
for any given set of data by mathematical programming techniques.

Actually, the inequality ($\leq$) restriction in (4) can without loss of
generality be replaced by equality. A proof is given in the appendix for
situations where at least two or three of the g_m are nonzero.

The writer is indebted to Martha Hamilton who developed the computer
program to carry out the required minimizations and maximizations numerically
for various sets of data. The program implements a sequential unconstrained
minimization technique (SUMT) of Fiacco and McCormick (1968, chapter 4).
The constraint on X^2 was handled by use of a penalty function; other
constraints were dealt with by simpler means. The unconstrained minimiza-
tions required for SUMT were carried out by a program developed by
Jöreskog (1967, section 8) and modified by Hamilton, implementing the
Fletcher-Powell-Davidon (1963) method.

All computer runs were made in double precision on an IBM 360/65.
As a check, each of the 44 extremization problems dealt with in Table 1
was run with two different starting points, one of which was completely
random within the limitations $0 \leq g_m$ and $0 \leq \zeta_m \leq 1$ ($m = 0,1,\ldots,M$)
and $\Sigma g_m = 1$. When $N = 12{,}990$, the agreement between the $\mu_{\alpha x}$ or $\bar{\mu}_{\alpha x}$
reached from two different starting points was to at least four decimal

12

places; when $N = 130$, the agreement was to at least three decimal places. This suggests that the intervals obtained represent global, not just local, maxima and minima, at least to a three-decimal-place approximation. Hundreds of other checks were made to be sure that small changes in $g_{\alpha x}(\zeta)$ or in $\bar{g}_{\alpha x}(\zeta)$ would not give more extreme $\mu_{\alpha x}$ or $\bar{\mu}_{\alpha x}$, respectively. All such checks were satisfied.

5. Numerical Results

The procedure described was applied to the real data shown in the second column of Table 1. This column shows the frequency distribution of $N = 12,990$ independent observations (actually, test scores of 12,990 students--their number of correct answers on a psychological test composed of $n = 20$ questions). A separate study (see below) shows that this distribution is compatible with the mathematical model given by (1) and (2).

The fifth column shows interval estimates of $\mu_{z|x}$, obtained for these data by the method outlined, with $\alpha = .05$. These empirical Bayes intervals, unlike ordinary interval estimates, are wider at extreme values of x than at middle values. In the middle, the intervals shown for $N = 12,990$ are happily short.

Although it is not obvious from a look at the table, no straight-line regression of ζ on x can be fitted inside the intervals shown for $N = 12,990$. The indicated nonlinearity is not rigorously demonstrated by the methods described here. However, linearity of regression implies

13

Table 1. Observed frequency distributions $f(x)$ and corresponding interval estimates ($\alpha = .05$) for the regression of Z on x .

x	$Nf(x)$		$\hat{\mu}_{Z|x}$	Interval Estimate of $\mu_{Z|x}$	
	$N = 12,990$	$N = 130$		$N = 12,990$	$N = 130$
20	63	2	.898	.852-.952	.690-1.000
19	141	2	.863		
18	220	2	.823	.780-.846	.615-.943
17	319	2	.773		
16	424	6	.716	.690-.743	.573-.855
15	622	4	.663		
14	776	8	.619	.605-.646	.510-.749
13	1001	4	.583		
12	1203	9	.553	.534-.564	.440-.636
11	1443	19	.526		
10	1550	13	.500	.485-.511	.404-.556
9	1409	12	.475	.461-.487	
8	1235	13	.451	.436-.463	.363-.491
7	1052	18	.426		
6	696	8	.402	.387-.420	.305-.491
5	471	4	.379		
4	226	3	.356	.340-.399	.199-.491
3	98	1	.334		
2	27	0	.314	.233-.383	.049-.491
1	12	0	.296	.107-.380	
0	2	0	.280	.010-.374	.000-.491
	12,990	130			

14

that $\phi_G(x)$ is a negative hypergeometric distribution (Lord & Novick, 1968, section 23.6); thus linearity can be tested by determining whether $f(x)$ can be considered a random sample from such a distribution.

In order to investigate the effects of sample size, a random sample of 130 observations was drawn from the 12,990. The resulting $f(x)$ is shown in the table along with the corresponding interval estimates of $\mu_{z|x}$. The intervals are of course much wider than for $N = 12,990$, but fortunately not 10 times as wide. (Grouping of frequencies was used in the calculations where necessary so that the denominator of (4) should never be less than 1.)

The fourth column of the table is included as a partial check on the validity of the intervals obtained. This column shows the regression of ζ on x corresponding to a certain $G(\zeta)$ which was found (in a separate study) to provide a good fit to the $f(x)$ in column two, the obtained chi square between $\phi_G(x)$ and $f(x)$ being near the 50th percentile of the tabled distribution of chi square for 20 degrees of freedom. It is pleasant to find that these regression values all lie well within the interval estimates shown in column five.

References

Copas, J. B. Compound decisions and empirical Bayes. Journal of the Royal
Statistical Society, Series B, 1969, 31, 397-425.

Fiacco, A. V. & McCormick, G. P. Nonlinear programming: Sequential
unconstrained minimization techniques. New York: Wiley, 1968.

Fletcher, R. & Powell, M. J. D. A rapidly convergent descent method for
minimization. Computer Journal, 1963, 2, 163-168.

Griffin, B. S. & Krutchkoff, G. Optimal linear estimators: an empirical
Bayes version with application to the binomial distribution.
Biometrika, 1971, 58, 195-201.

Jöreskog, K. G. Some contributions to maximum likelihood factor analysis.
Psychometrika, 1967, 32, 443-482.

Karlin, S. & Shapley, L. S. Memoirs of the American Mathematical Society.
Number 12. Geometry of moment spaces. Providence: American
Mathematical Society, 1953.

Lord, F. M. & Novick, M. R. Statistical theories of mental test scores.
Reading, Mass.: Addison-Wesley, 1968.

Maritz, J. S. Smooth empirical Bayes estimation for one-parameter discrete
distributions. Biometrika, 1966, 53, 417-429.

Possé, C. Sur quelques applications des fractions continues algébriques.
St. Pétersbourg, Russie: L'Académie Impériale des Sciences, 1886.

Robbins, H. An empirical Bayes approach to statistics. In J. Neyman (Ed.),
Proceedings of the Third Berkeley Symposium on Mathematical Statistics
and Probability. Berkeley: University of California Press, 1956.
Pp. 157-164.

Skellam, J. G. A probability distribution derived from the binomial distribution by regarding the probability of success as variable between sets of trials. Journal of the Royal Statistical Society, Series B, 1948, 10, 257-261.

Appendix

It will be shown here that $g_{\alpha x}(\zeta)$, the $G(\zeta)$ that minimizes (6) or (8) subject to (4) and (9) must lie on the boundary where $\chi_G^2 = \chi_{1-\alpha}^2$. To avoid tediousness, the proofs are limited to situations where at least two or three (as convenient) of the ζ_m are distinct and have nonzero frequency.

It will be shown first that $\mu_{Z|x}$, considered as a function of the g_m and the ζ_m, has no <u>unrestricted</u> minimum satisfying (9). This is proved by showing that if we are given any set with $K \geq 2$ of g_m and ζ_m satisfying (9), we can reduce the $\mu_{Z|x}$ computed from (8) by a change in the g_m and ζ_m that does not violate (9). Thus if a <u>restricted</u> minimum with $K \geq 2$ is found by the methods of section 4, it must lie on the only remaining boundary ($\chi_G^2 = \chi_{1-\alpha}^2$), imposed by restriction (4), since otherwise $\mu_{Z|x}$ could still be reduced by the methods outlined. We will treat explicitly only the case where x is even. Thus $\zeta_0 = 0$ by Markov's theorem. (Similar proofs will apply when x is odd.) Let $\zeta_1 < \ldots < \zeta_K$ be the $K \geq 2$ distinct nonzero values at which $g_{\alpha x}(\zeta)$ has nonzero frequency.

All possibilities will now be covered by the following four cases.

<u>Case 1</u>: $g_0 > 0$, $0 < x < n$.

Consider the derivative of (8) with respect to ζ_i:

$$\frac{\partial \mu_{Z|x}}{\partial \zeta_i} \propto g_i \zeta_i^x (1 - \zeta_i)^{n-x-1}[x + 1 - (n + 1)\zeta_i] \sum_{m=0}^{K} g_m \zeta_m^x (1 - \zeta_m)^{n-x}$$

$$- g_i \zeta_i^{x-1}(1 - \zeta_i)^{n-x-1}(x - n\zeta_i) \sum_{m=0}^{K} g_m \zeta_m^{x+1}(1 - \zeta_m)^{n-x}$$

18

$$\propto g_i \zeta_i^{x-1}(1 - \zeta_i)^{n-x-1}$$

$$\cdot \sum_{m=0}^{K} g_m \zeta_m^x (1 - \zeta_m)^{n-x}[\zeta_i\{x + 1 - (n + 1)\zeta_i\} - \zeta_m(x - n\zeta_i)]$$

$$\propto g_i \zeta_i^{x-1}(1 - \zeta_i)^{n-x-1}$$

$$\cdot \sum_{m=0}^{K} g_m \zeta_m^x (1 - \zeta_m)^{n-x}[\zeta_i(1 - \zeta_i) + (\zeta_i - \zeta_m)(x - n\zeta_i)] \quad . \quad (A1)$$

Take $i = 0$ and suppose for the moment that $K \geq 2$ and that ζ_0 $(= 0)$ is replaced by a very small positive quantity. The first term in the brackets can be neglected. The second term in the brackets is zero when $m = 0$ but is negative otherwise. Since all quantities outside the brackets are nonnegative and some, at least, are positive, the derivative (A1) will be negative. Thus $\mu_{Z|x}$ can be reduced by replacing ζ_0 by a small positive quantity (this can also be seen intuitively).

Case 2: $g_0 = 0$, $0 < x < n$.

Consider the effect on $\mu_{Z|x}$ of a small change in g_i and a small compensating change in g_K, holding all other g_m fixed. Treating g_K as a function of g_i defined by the equation

$$g_K = 1 - \sum_{m=0}^{K-1} g_m \qquad (A2)$$

and using the formula

$$\frac{d\mu_{Z|x}}{dg_i} = \frac{\partial \mu_{Z|x}}{\partial g_i} + \frac{\partial \mu_{Z|x}}{\partial g_K} \frac{dg_K}{dg_i} \quad , \qquad (A3)$$

19

we find from (8) that when the g_m $(m \neq i,K)$ are fixed,

$$\frac{d\mu_{Z|x}}{dg_i} \propto \sum_{m=0}^{K} g_m \zeta_m^x (1 - \zeta_m)^{n-x}[\zeta_i^{x+1}(1 - \zeta_i)^{n-x} - \zeta_K^{x+1}(1 - \zeta_K)^{n-x}]$$

$$- \sum_{m=0}^{K} g_m \zeta_m^{x+1}(1 - \zeta_m)^{n-x}[\zeta_i^x(1 - \zeta_i)^{n-x} - \zeta_K^x(1 - \zeta_K)^{n-x}]$$

$$\propto \sum_{m=0}^{K} g_m \zeta_m^x (1 - \zeta_m)^{n-x}[\zeta_i^{x+1}(1 - \zeta_i)^{n-x}$$

$$- \zeta_K^{x+1}(1 - \zeta_K)^{n-x} - \zeta_m \zeta_i^x(1 - \zeta_i)^{n-x} + \zeta_m \zeta_K^x(1 - \zeta_K)^{n-x}]$$

$$\propto \sum_{m=1}^{K} g_m \zeta_m^x (1 - \zeta_m)^{n-x}[\zeta_i^x(1 - \zeta_i)^{n-x}(\zeta_i - \zeta_m)$$

$$+ \zeta_K^x(1 - \zeta_K)^{n-x}(\zeta_m - \zeta_K)] \quad . \tag{A4}$$

The last expression in (A4) makes use of the fact that since $g_0 = 0$ for Case 2, any summation in (A4) can be written either including or excluding $m = 0$.

The second term in the brackets is never positive. Now take $i = 1$. The first term in brackets is now zero when $m = 1$ and negative otherwise. All quantities outside the brackets are nonnegative; if either $\zeta_K < 1$, or if $K \geq 3$, then some of these quantities are positive. If so, the derivative of (A4) is necessarily negative for $i = 1$. Thus $\mu_{Z|x}$ can be reduced by shifting some frequency from ζ_1 to ζ_K .

(If $\zeta_K = 1$ and $K = 2$, all terms under the summation in (A4) are zero. This special case can be dealt with by using (A1) again with $i = 1$. Since $g_0 = 0$, $1 - \zeta_K = 0$, and $\zeta_i - \zeta_1 = 0$, the last expression in

(A1) simplifies to $d\mu_{Z|x}/d\zeta_1 \propto g_1^2 \zeta_1^{2x}(1 - \zeta_1)^{2n-2x}$. This derivative of $\mu_{Z|x}$ is necessarily positive for $x < n$. Thus $\mu_{Z|x}$ can be reduced by decreasing ζ_1 .)

Case 3: $x = 0$

In this case (8) becomes

$$\mu_{Z|x} = \frac{\sum\limits_{m=0}^{M} g_m \zeta_m (1 - \zeta_m)^n}{\sum\limits_{m=0}^{M} g_m (1 - \zeta_m)^n} \quad .$$

Using (A2) and (A3) as before, but with $i = 0$, we find, since $\zeta_0 = 0$, that

$$\frac{d\mu_{Z|x}}{dg_0} \propto \sum_{m=0}^{M} g_m (1 - \zeta_m)^n [- \zeta_K (1 - \zeta_K)^n]$$

$$- \sum_{m=0}^{M} g_m \zeta_m (1 - \zeta_m)^n [(1 - \zeta_0)^n - (1 - \zeta_K)^n]$$

$$\propto \sum_{m=0}^{M} g_m (1 - \zeta_m)^n [- \zeta_K (1 - \zeta_K)^n - \zeta_m + \zeta_m (1 - \zeta_K)^n]$$

$$\propto \sum_{m=0}^{M} g_m (1 - \zeta_m)^n [- \zeta_m - (\zeta_K - \zeta_m)(1 - \zeta_K)^n] \quad . \qquad (A5)$$

Since (A5) is always negative, an increase in g_0 together with a corresponding decrease in g_K will reduce $\mu_{Z|x}$ in Case 3.

Case 4: $x = n$.

In this case

$$\mu_{Z|x} = \frac{\sum\limits_{m=1}^{M} g_m \zeta_m^{n+1}}{\sum\limits_{m=1}^{M} g_m \zeta_m^{n}} \quad .$$

Using (A2) and (A3) as before with $i = 0$, we have

$$\frac{d\mu_{Z|x}}{dg_0} \propto \sum_{m=1}^{M} g_m \zeta_m^{n}(-\zeta_K^{n+1}) + \sum_{m=1}^{M} g_m \zeta_m^{n+1}(\zeta_K^{n})$$

$$\propto \zeta_K^{n} \sum_{m=1}^{M} g_m \zeta_m^{n}(\zeta_m - \zeta_K) \quad . \tag{A6}$$

Since (A6) is always negative, an increase in g_0 together with a compensating decrease in g_K will reduce $\mu_{Z|x}$ in Case 4.

The four cases listed are exhaustive, provided $K \geq 2$. Similar proofs could be written to eliminate this proviso. Consequently, a minimum for $\mu_{Z|x}$ cannot occur except on the boundary established by (4).

22

4 Director, Personnel and Training
   Research Programs
   Office of Naval Research
   Arlington, Va. 22217

1 Director
   ONR Branch Office
   495 Summer Street
   Boston, Mass. 02210

1 Director
   ONR Branch Office
   1030 East Green Street
   Pasadena, Cal. 91101

1 Director
   ONR Branch Office
   536 South Clark Street
   Chicago, Ill. 60605

1 Office of Naval Research
   Area Office
   207 West 24th Street
   New York, N. Y. 10011

1 Director, Information Systems Program
   Office of Naval Research (Code 437)
   Arlington, Va. 22217

6 Director
   Naval Research Laboratory
   Washington, D. C. 20390
   Attn: Library, Code 2029 (ONRL)

6 Director
   Naval Research Laboratory
   Washington, D. C. 20390
   Attn: Technical Information Division

12 Defense Documentation Center
    Cameron Station, Building 5
    5010 Duke Street
    Alexandria, Va. 22314

1 Behavioral Sciences Department
   Naval Medical Research Institute
   National Naval Medical Center
   Bethesda, Md. 20014

1 Chief
   Bureau of Medicine and Surgery
   Code 513
   Washington, D. C. 20390

1 Chief
   Bureau of Medicine and Surgery
   Research Division (Code 713)
   Department of the Navy
   Washington, D. C. 20390

1 Commanding Officer
   Naval Medical Neuropsychiatric
     Research Unit
   San Diego, Cal. 92152

1 Director
   Education and Training Sciences
     Department
   Naval Medical Research Institute
   National Naval Medical Center
   Building 142
   Bethesda, Md. 20014

1 Technical Reference Library
   Naval Medical Research Institute
   National Naval Medical Center
   Bethesda, Md. 20014

1 Mr. S. Friedman
   Special Assistant for Research
     & Studies
   OASN (M&RA)
   The Pentagon, Room 4E794
   Washington, D. C. 20350

1 Chief
   Naval Air Technical Training
   Naval Air Station
   Memphis, Tenn. 38115

1 Naval Air Systems Command
     (AIR 5313A)
   Washington, D. C. 20360

1 Chief of Naval Operations (Op-98)
   Department of the Navy
   Washington, D. C. 20350
   Attn: Dr. J. J. Collins

3 Technical Director
   Personnel Research Division
   Bureau of Naval Personnel
   Washington, D. C. 20370

3 Technical Library (Pers-11B)
   Bureau of Naval Personnel
   Department of the Navy
   Washington, D. C. 20360

1 Training Research Program Mngr
   Bureau of Naval Personnel (Pers-A321)
   Washington, D. C. 20370

3 Technical Director
   Naval Personnel Research and
     Development Laboratory
   Washington Navy Yard, Building 200
   Washington, D. C. 20390

3 Commanding Officer
   Naval Personnel and Training
     Research Laboratory
   San Diego, Cal. 92152

1 Chairman
   Behavioral Science Department
   Naval Command and Management Division
   U. S. Naval Academy
   Luce Hall
   Annapolis, Md. 21402

1 Superintendent
   Naval Postgraduate School
   Monterey, Cal. 93940
   Attn: Library (Code 2124)

1 Commanding Officer
   Service School Command
   U. S. Naval Training Center
   San Diego, Cal. 92133

1 Research Director, Code 06
   Research and Evaluation Department
   U. S. Naval Examining Center
   Building 2711--Green Bay Area
   Great Lakes, Ill. 60088
   Attn: C. S. Winiewicz

1   Commander
    Submarine Development Group Two
    Fleet Post Office
    New York, N. Y.   09501

1   Mr. George N. Graine
    Naval Ship Systems Command (SHIP 03H)
    Department of the Navy
    Washington, D. C.   20360

1   Head, Personnel Measurement Staff
    Capital Area Personnel Service Office
    Ballston Tower #2, Room 1204
    801 N. Randolph Street
    Arlington, Va.   22203

1   Mr. Leo Mason
    Center for Naval Analyses
    1400 Wilson Blvd.
    Arlington, Va.   22209

1   Dr. Robert Lockman
    Center for Naval Analysis
    1400 Wilson Blvd.
    Arlington, Va.   22209

1   Dr. A. L. Slafkosky
    Scientific Advisor (Code AX)
    Commandant of the Marine Corps
    Washington, D. C.   20380

1   Dr. James J. Regan, Code 55
    Naval Training Device Center
    Orlando, Fla.   32813

1   Behavioral Sciences Division
    Office of Chief of Research and
        Development
    Department of the Army
    Washington, D. C.   20310

1   U. S. Army Behavior and Systems
        Research Laboratory
    Commonwealth Building, Room 239
    1320 Wilson Boulevard
    Arlington, Va.   22209

1   Director of Research
    U. S. Army Armor Human Research
        Unit
    Attn: Library
    Bldg. 2422 Morande Street
    Fort Knox, Ky.   40121

1   Commanding Officer
    Attn: LTC Cosgrove
    USA CDC PASA
    Ft. Harrison, Ind.   46249

1   Dr. Vincent Cieri
    Education Advisor
    U. S. Army Signal Center and School
    Fort Monmouth, N. J.   07703

1   Mr. Harold A. Schultz
    Educational Advisor-ATIT-E
    CONARC
    Fort Monroe, Va.   23351

1   Director
    Behavioral Sciences Laboratory
    U. S. Army Research Institute of
        Environmental Medicine
    Natick, Mass.   01760

1   Division of Neuropsychiatry
    Walter Reed Army Institute of
        Research
    Walter Reed Army Medical Center
    Washington, D. C.   20012

1   Dr. George S. Harker, Director
    Experimental Psychology Division
    U. S. Army Medical Research Laboratory
    Fort Knox, Ky.   40121

1   AFHRL (TR/Dr. G. A. Eckstrand)
    Wright-Patterson Air Force Base
    Ohio   45433

1   AFHRL (TRT/Dr. Ross L. Morgan)
    Wright-Patterson Air Force Base
    Ohio   45433

1   AFHRL (TRR/Dr. Melvin T. Snyder)
    Air Force Human Resources Lab.
    Wright-Patterson Air Force Base
    Ohio   45433

1   AFSOR (NL)
    1400 Wilson Boulevard
    Arlington, Va.   22209

1   Lt. Col. Robert R. Gerry, USAF
    Chief, Instructional Technology
        Programs
    Resources & Technology Division
    (DPTBD DCS/P)
    The Pentagon (Room 4C244)
    Washington, D. C.   20330

1   Headquarters, U. S. Air Force
    Chief, Personnel Research and
        Analysis Division (AFlDPXY)
    Washington, D. C.   20330

1   Personnel Research Division (AFHRL)
    Lackland Air Force Base
    San Antonio, Tex.   78236

1   Commandant
    U. S. Air Force School of
        Aerospace Medicine
    Attn: Aeromedical Library
    Brooks AFB, Tex.   78235

1   Headquarters, Electronics Systems
        Division
    Attn: Dr. Sylvia Mayer/MCDS
    L. G. Hanscom Field
    Bedford, Mass.   01730

1   Lt. Col. Austin W. Kibler
    Director, Behavioral Sciences (Acting)
    Advanced Research Projects Agency,
        DDR&E
    1400 Wilson Boulevard
    Arlington, Va.   22209

1   Director of Manpower Research
    OASD (M&RA) (M&RU)
    Room 3D960
    The Pentagon
    Washington, D. C.   20330

1   Mr. Joseph J. Cowan, Chief
    Psychological Research Branch (P-1)
    U. S. Coast Guard Headquarters
    400 Seventh Street, S. W.
    Washington, D. C.   20591

24

1 Dr. Alvin E. Goins, Chief
   Personality and Cognition Research
     Section
   Behavioral Sciences Research Branch
   National Institute of Mental Health
   5454 Wisconsin Ave., Room 10A01
   Bethesda, Md.   20014

1 Dr. Andrew R. Molnar
   Computer Innovation in Education
     Section
   Office of Computing Activities
   National Science Foundation
   Washington, D. C.   20550

1 Dr. Richard C. Atkinson
   Department of Psychology
   Stanford University
   Stanford, Cal.   94305

1 Dr. Bernard M. Bass
   University of Rochester
   Management Research Center
   Rochester, N. Y.   14627

1 Dr. Lee R. Beach
   Department of Psychology
   University of Washington
   Seattle, Washington   98105

1 Dr. Mats Bjorkman
   University of Umea
   Department of Psychology
   Umea 6, SWEDEN

1 Dr. Jaime R. Carbonell
   Bolt, Beranek & Newman, Inc.
   50 Moulton Street
   Cambridge, Mass.   02138

1 Dr. Lee J. Cronbach
   School of Education
   Stanford University
   Stanford, Cal.   94305

1 Dr. Marvin D. Dunnette
   University of Minnesota
   Department of Psychology
   Elliot Hall
   Minneapolis, Minn.   55455

1 Division of Analytical Studies
     and Services
   Educational Testing Service
   Rosedale Road
   Princeton, N. J.   08540

1 Dr. Robert Glaser
   Learning Research and Development
     Center
   University of Pittsburgh
   Pittsburgh, Pa.   15213

1 Dr. Albert S. Glickman
   American Institutes for Research
   8555 Sixteenth Street
   Silver Spring, Md.   20910

1 Dr. Bert Green
   Department of Psychology
   Johns Hopkins University
   Baltimore, Md.   21218

1 Dr. Harold Gulliksen
   Department of Psychology
   Princeton University
   Princeton, N. J.   08540

1 Dr. Duncan N. Hansen
   Center for Computer Assisted Instruction
   Florida State University
   Tallahassee, Fla.   32306

1 Dr. Richard S. Hatch
   Decision Systems Associates, Inc.
   11428 Rockville Pike
   Rockville, Md.   20852

1 Dr. M. D. Havron
   Human Sciences Research, Inc.
   Westgate Industrial Park
   7710 Old Springhouse Road
   McLean, Va.   22101

1 Dr. Albert E. Hickey
   Entelek, Incorporated
   42 Pleasant Street
   Newburyport, Mass.   01950

1 Human Resources Research Organization
   Library
   300 North Washington Street
   Alexandria, Va.   22314

1 Human Resources Research Organization
   Division #3
   Post Office Box 5787
   Presidio of Monterey, Cal.   93940

1 Human Resources Research Organization
   Division #4, Infantry
   Post Office Box 2086
   Fort Benning, Ga.   31905

1 Human Resources Research Organization
   Division #5, Air Defense
   Post Office Box 6021
   Fort Bliss, Tex.   77916

1 Human Resources Research Organization
   Division #6, Aviation (Library)
   Post Office Box 428
   Fort Rucker, Ala.   36360

1 Dr. Robert R. Mackie
   Human Factors Research, Inc.
   Santa Barbara Research Park
   6780 Cortona Drive
   Goleta, Cal.   93017

1 Dr. Stanley M. Nealy
   Department of Psychology
   Colorado State University
   Fort Collins, Colo.   80521

1 Mr. Luigi Petrullo
   2431 North Edgewood Street
   Arlington, Va.   22207

1 Psychological Abstracts
   American Psychological Association
   1200 Seventeenth Street, N. W.
   Washington, D. C.   20036

1   Dr. Diane M. Ramsey-Klee
    R-K Research & System Design
    3947 Ridgemont Drive
    Malibu, Cal.   90265

1   Dr. Joseph W. Rigney
    Behavioral Technology Laboratories
    University of Southern California
    University Park
    Los Angeles, Cal.   90007

1   Dr. George E. Rowland
    Rowland and Company, Inc.
    Post Office Box 61
    Haddonfield, N. J.   08033

1   Dr. Arthur I. Siegel
    Applied Psychological Services
    Science Center
    404 East Lancaster Avenue
    Wayne, Pa.   19087

1   Dr. Ledyard R Tucker
    University of Illinois
    Psychology Building
    Urbana, Ill.   61820